Condition Monitoring with Incomplete Observations

Jonas Vlasselaer[1], **Wannes Meert**[1] , **Rocco Langone**[2] and **Luc De Raedt**[1]

Abstract.
We introduce an approach for predicting the behaviour of a machine during a production cycle. Typical data analysis methods assume that continuous behaviour is (fully) observed. This assumption is unrealistic as monitored machines are often interrupted and restarted at irregular points in time. We study the resulting problem, propose a solution and report on a use-case in wire drawing.

1 PROBLEM SETTING

Condition monitoring allows a company to analyze the behaviour of their machinery and predict time-to-maintenance or time-to-failure. The latter is useful as unexpected down-time caused by failures not only decreases production capacity but also increases maintenance cost. In current practice, the operator often tries to avoid failures by preventively shutting down the machine or changing the production settings when an alarming trend in the behaviour is observed.

The process of machine monitoring typically results in a number of observed variables like continuous sensor readings, configuration settings and environmental factors. Based on these, we want to predict future values of one or more target variables (e.g. temperature, vibration) because it allows to choose more optimal actions. If the target variable is completely observed from its minimal to its maximal value for a variety of settings, we can apply a regression model (e.g. SVM) to model the behaviour (see for example Fig. 1.d). Unfortunately, this behaviour is not always fully observed because the machines are interrupted by external interventions (e.g. batches, shifts, change of raw material).

In the given environment, the challenge is to correctly combine those fragments that are generated by the same set of relevant, possibly unobserved, variables. Once the generative model which explains the complete behaviour is reconstructed, it can be used to (1) predict future behaviour for a given configuration of settings, (2) detect anomalies which require further investigation, and (3) perform a data quality analysis to know whether the machine behaviour can be fully explained by the given set of observations.

This type of data is not only typical when monitoring machines, but also occurs in other fields like, for example, health monitoring. Performing exercises has an impact on the heart rate of a patient but this trend is frequently interrupted by a change of exercise or a pause and can be influenced by environmental factors like wind.

1.1 Wire Drawing

The use case we present handles on wire drawing and is in collaboration with a wire processing company. Their aim is to maximize production capacity based on temperature condition monitoring. The machines of interest transform a given steel wire to a product that meets with the customers specifications. A production cycle consists of (1) mounting a spool of wire at the input, (2) process the entire spool and (3) remove the finished product. This process is repeated multiple times a day but can be interrupted due to failures, e.g. a wire rupture, which causes a short pause in step 2. When active, energy dissipation causes the machines to heat while a pause, due to a failure or mounting a new spool, causes it to cool down.

Experience has shown that certain components in the machines, especially the bearings, tend to degrade faster when high temperatures are measured during operation. Because this early failing of components results in higher costs, the operators prevent high temperatures by shutting down a machine once a threshold is reached and only restart production after it is sufficiently cooled down. This strategy, however, is suboptimal to maximize production capacity as a different machine setting might avoid the machine to overheat without the need to interrupt the production cycle. Hence the data analysis task in this use-case is twofold. The first task is to build a model that estimates the temperature evolution based on a machine's settings and sensor readings. Secondly, this model is used to plan for the optimal combination of orders and settings. In this paper, we focus on the former task.

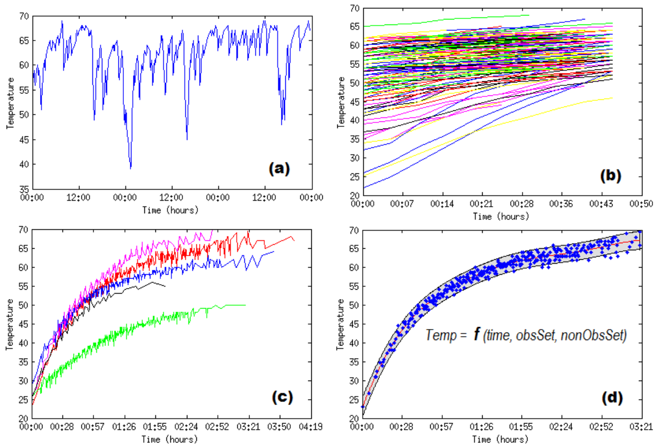

Figure 1. Reconstructing the heating behaviour. (a) Measured temperature, (b) heating fragments, (c) heating profiles and (d) predictive model.

1.2 The Dataset

The dataset contains machine settings and sensor measurements of a collection of machines for a period of about two years. The logged settings include, for example, rotations per minutes and size of the spool. The measurements include the temperature at the input and output of the machine (both need to remain under a given threshold). Additionally, the status of the machine is registered (e.g. running or not, wire rupture, spool finished).

---

[1] KU Leuven, CS, Belgium, email: name.surname@cs.kuleuven.be
[2] KU Leuven, ESAT, Belgium, email: name.surname@esat.kuleuven.be

The machines are stopped frequently, after each production cycle or when a failure occurs. Because the down-time is unpredictable and irregular, the machine often restarts at different temperatures and conditions and, as such, the dataset does not contain full heating curves. It thus contains a latent variable that corresponds to the time it would take for the machine to reach the current temperature under the current settings. Before a typical forecasting or regression algorithm can be used, the latent variable needs to be quantified.

2 TRAINING THE MODEL

2.1 Data Pre-Processing

The available temperature data is a continuous stream of measurements as shown in Fig. 1.a. The relevant unit to model the heating behaviour, however, are the *fragments* generated while the machine is active. These fragments, as shown in 1.b, are extracted from the dataset by analyzing the status of the machine and are saved together with the machine settings. We retain all heating fragments with more than five measurements, irrespective of the reason why the machine is stopped. A similar set of fragments can be extracted while the machine is paused and cools down. Without loss of generality, we only discuss the approach for heating fragments.

2.2 Data Alignment

The essential step is to combine multiple heating fragments to reconstruct the complete heating behaviour. Since there is no guarantee that we observe all variables that influence the behaviour or that all variables are relevant, we cannot naively combine fragments obtained with identical settings. Therefore, we use a hierarchical clustering approach to cluster the fragments based on similarity with Dynamic Time Warping (DTW) [1] as distance function. The different clusters are expected to correspond to different machine settings.

We use DTW to find sufficiently overlapping fragments that behave similar while tolerating for subtle delays in heating. It calculates the distance between two heating fragments not only based on the difference in temperature but also on a possible shift at one or more points in time. It serves as an estimation of the latent variable. The hierarchical clustering results in a set of clusters based on the maximal distance between a fragment and a cluster's prototype. We call the combination of all heating fragments in one cluster a *heating profile*. Five of these profiles are depicted in Fig. 1.c.

2.3 Regression Model

The obtained heating profiles can be used to learn a regression model to predict changes in temperature. For this purpose, we use Least Squares Support Vector Machines (LS-SVMs) [2]. Support Vector Machines (SVMs) are a well known classification and regression method which perform their task in a high-dimensional kernel-induced feature space. LS-SVMs differ from SVMs for the presence of a quadratic loss function in the primal problem and equality instead of inequality constraints.

Since LS-SVMs use a quadratic loss function, outliers can have a large influence on the estimated model. To improve the robustness, a weighted LS-SVM technique is used. The weighted LS-SVM are especially useful in this setting as we do not have much measurements at higher temperatures and as such our profile might contain outliers at this region. Additionally, LS-SVM also returns confidence bounds on the prediction. We show the model for one of the obtained heating profiles in figure 1.d.

3 APPLYING THE LEARNED MODEL

The result of training is a two-layer model: (1) a set of heating profiles and (2) a predictive model for each profile. To know when to apply which model we need to know when they are applicable. To achieve this we try to explain the different profiles by looking at the machine configuration (and possibly the environment). This was realized using decision trees with the profile as the target variable. As a result, the complete model we obtain is a tree with, close to the top, the machine settings which influence the machine behaviour and at each leaf a regression model trained on all the fragments in the leaf.

Analysis of the obtained decision tree trained for each machine individually revealed that speed is the most important machine setting. Therefore, we can approximate our model by first splitting the available data on the different set speeds (about five fixed values) and learn different regression models for each of these speeds. In this way we actually have a one-layer decision tree where we ignore variables other than speed and temperature. For this model, we also assume we deal with a Markov process where, given the initial temperature and set speed, we do not need any information of the past of the machine.

Additionally, the decision tree can be used for two other tasks. Firstly, to infer whether enough machine settings and environment characteristics were monitored to explain different profiles. If two profiles end up in the same leaf of a tree, an extra variable needs to be monitored to be able to distinguish between them. Secondly, to observe a variation over time. If the machine is not working as expected or an alteration was performed, the learned model will report a divergence from expected behaviour. This is useful for anomaly detection.

4 EXPERIMENTS

We briefly present some of the experimental results obtained with the methods explained in this paper. (1) The predictive models obtained after splitting the data on different set speeds allows to estimate the future temperature of our test-machine with an R^2 coefficient of 0.93. (2) Hierarchical clustering of all fragments, obtained from all machines, results in 16 relevant heating profiles. They could be explained by the machine settings with an accuracy of 75 %. As the different heating profiles cannot be completely explained by the machine settings, the dataset seems to miss some relevant settings or context. (3) After we added time-information as an extra variable, our model showed there is a clear shift in behavior of one machine at one specific moment in time. The company itself was surprised by this information and now investigates which missing information could be the cause for this shift such that it also can be logged and added to the dataset.

Acknowledgements. This work was supported by the IWT-SBO POM2 Project (100031). We thank the Flanders Mechatronics Technology Centre for their support and ensuring access to the data.

REFERENCES

[1] H. Sakoe and S Chiba, 'Dynamic programming algorithm optimization for spoken word recognition', *IEEE Trans. on Acoust., Speech, and Signal Process.*, **26**, 43–49, (1978).

[2] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002.